DOT/FAA/AR-95/16, I

Office of Aviation Research
Washington, D.C. 20591

Test and Evaluation Report for the Explosive Device Detection Baseline Study

J. L. Fobes, Ph.D.
D. Michael McAnulty, Ph.D.
Eric Neiderman, Ph.D.
Brenda A. Klock
J. Michael Barrientos

Aviation Security Human
Factors Program, AAR-510
FAA Technical Center
Atlantic City International Airport, NJ 08405

December 1995

Final

WARNING NOTICE: This document is a record subject to the provisions of 14 CFR 191.1 et seq. Release of information contained herein is prohibited without the express written approval of the Associate Administrator for Civil Aviation Security, Federal Aviation Administration, or his or her designee.

U.S. Department of Transportation
Federal Aviation Administration

NOTICE

Technical Report Documentation Page

1. Report No. DOT/FAA/AR-95/16	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Test and Evaluation Plan for the Explosive Device Detection Baseline Study		5. Report Date December 1995 6. Performing Organization Code AAR-510
7. Author(s) J. L. Fobes, Ph.D., D. Michael McAnulty, Ph.D., E. Neiderman, Ph.D., J. M. Barrientos, and B. A. Klock		8. Performing Organization Report No.
9. Performing Organization Name and Address U.S. Department of Transportation Federal Aviation Administration FAA Technical Center Atlantic City International Airport, NJ 08405		10. Work Unit No. (TRAIS) 11. Contract or Grant No. DTFA03-89-C-00043
12. Sponsoring Agency Name and Address U.S. Department of Transportation Federal Aviation Administration FAA Headquarters 800 Independence Ave., S.W. Washington, DC 20591		13. Type of Report and Period Covered Final 14. Sponsoring Agency Code ACS-1

15. Supplementary Notes

Draft prepared by J. Janowitz, Ed.D., R. L. Malone, J. Brasher, D. Fischer, and B. Schultz of Galaxy Scientific Corporation.

16. Abstract

This document is the Test and Evaluation Report evaluating the ability of airport security personnel to detect improvised explosive devices (IEDs) in carry-on passenger bags. The test and evaluation focused on determining the baseline performance levels as set forth in the Critical Operational Issues and Criteria.

The test was conducted at all of the 19 U.S. Category X airports. The screeners' ability to locate IEDs, as assessed by their probability of detection and signal detection theory measure of d′, was similar at most CAT X airports. Some differences in performance were found at the extremes. Similarly, no important differences in performance were found among the security companies represented.

Training should be designed to increase screener d′ and P_d while sustaining or lowering screener P_{fa} by improving the ability of screeners to discriminate IEDs from normal objects in passenger baggage.

17. Key Words X-ray, X-ray screening, IED, improvised explosive device, IED detection, signal detection, Human Factors Engineering	18. Distribution Statement This document is a record subject to the provisions of 14 CFR 191.1 et seq. Release of information contained herein is prohibited without the express written approval of the Associate Administrator for Civil Aviation Security, Federal Aviation Administration, or his or her designee.

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	37	

Form DOT F1700.7 (8-72)Reproduction of completed page authorized

PREFACE

This test report was developed to support investigation of the Critical Operational Issues and Criteria set forth by the Federal Aviation Administration to assess the capability of airport baggage screeners to detect improvised explosive devices in carry-on passenger baggage using black/white X-ray images. The key FAA personnel supporting this testing effort were J. L. Fobes, Ph.D., Aviation Security Human Factors Program Manager and Engineering Research Psychologist for the Aviation Security Research and Development Division (AAR-510), D. Michael McAnulty, Ph.D., an Engineering Research Psychologist with the Aviation Simulation and Human Factors Division (ACT-500), Eric Neiderman, Ph.D., Aviation Security Human Factors Engineering Research Psychologist (AAR-510), Brenda A. Klock, Technical Specialist, and J. Michael Barrientos, Aerospace Engineer for the Aviation Security Research and Development Division (AAR-510).

TABLE OF CONTENTS

APPENDIXES

LIST OF TABLES

ACRONYMS AND ABBREVIATIONS

AAR	Aviation Security Research and Development Division
ACS	Office of the Associate Administrator for Civil Aviation Security
ANOVA	Analysis of Variance
$\underline{c}$	Operator Response Criterion
CAT X	Category X
CO	Carry-On
COIC	Critical Operational Issue and Criteria
CR	Correct Rejection
d'	d prime, Derived Operator Sensitivity
DOT	Department of Transportation
EDDB	Explosive Device Detection Baseline
FAA	Federal Aviation Administration
FAR	Federal Aviation Regulation
IED	Improvised Explosive Device
MOP	Measure of Performance
N_c	Number of Comparison Bag Images
N_{fa}	Number of False Alarms
N_h	Number of Hits
NS	Not So Sure
N_t	Number of Test Bag Images
P_d	Probability of Detection
P_{fa}	Probability of False Alarm
ROC	Receiver Operating Characteristic
SDT	Signal Detection Theory
SPEARS	Screener Proficiency Evaluation and Reporting System
T&E	Test and Evaluation
TEP	Test and Evaluation Plan
TER	Test and Evaluation Report
VS	Very Sure

Airport Codes

ATL	Atlanta William B. Hartsfield International Airport
BOS	General Edward Lawrence Logan International Airport
BWI	Baltimore-Washington International Airport
DCA	Washington National Airport
DFW	Dallas/Fort Worth International Airport
DIA	Denver International Airport
DTW	Detroit Metropolitan Wayne County Airport
HNL	Honolulu International Airport
IAD	Washington Dulles International Airport

IAH	Houston Intercontinental Airport
JFK	John F. Kennedy International Airport
LAX	Los Angeles International Airport
MCO	Orlando International Airport
MIA	Miami International Airport
ORD	O'Hare International Airport
SEA	Seattle-Tacoma International Airport
SFO	San Francisco International Airport
SJU	Luis Munoz Marin International Airport (San Juan)
STL	Lambert St. Louis International Airport

EXECUTIVE SUMMARY

The purpose of this Explosive Device Detection Baseline (EDDB) study (Fobes, McAnulty, & Klock, 1995a) was to evaluate the capability of airport baggage screeners to detect improvised explosive devices (IEDs) in carry-on passenger baggage using black/white computer-based X-ray image presentations. The results establish a baseline of current aviation security performance. As new screener training technology and procedures are developed, performance improvements can be compared against this baseline. The Aviation Security Human Factors Airport Demonstration Test and Evaluation Master Plan (Fobes, McAnulty, & Klock, 1995b) describes various techniques to increase screener sensitivity (d′) and the probability of detection (P_{d}), while sustaining or lowering the probability of a false alarm (P_{fa}).

IED detection testing was performed using computer-based, X-ray image presentation. Two hundred digitized bag images were shown to 521 screeners from the 19 Category X (CAT X) airports. Simulated IEDs, built from modular bomb sets, were present in 25 bag images. Screeners were requested to indicate which images contained IEDs.

Three issues were examined to determine (a) the baseline IED detection performance at each U.S. CAT X airport, (b) significant differences in IED detection among CAT X airports, and (c) significant differences in IED detection among airport security companies at CAT X airports.

Although these issues were resolved in part by examining the P_d, additional signal detection theory (SDT) measures were included. Relying solely on the P_d can give a distorted estimate of screener capability because the P_d and accompanying P_{fa} covary considerably as a function of screener decision strategy. The SDT measures are emphasized here to analyze the two aspects of detection performance: the sensitivity with which IEDs were discriminated (d′) and the decision criteria (c) used.

The screeners' ability to locate IEDs was similar at all the CAT X airports, with some differences found at the extremes. Sensitivity averaged a d′ of 1.29, and screeners at the Honolulu International Airport (d′ = 1.68) found significantly more bomb images than did those at 14 other airports. Screeners at the Los Angeles International Airport (d′ = 1.00) found significantly fewer bomb images than did those at 9 other airports. The P_d measure of sensitivity averaged 0.55 and did not significantly differ among airports. The airport P_{fa} scores averaged 0.17 and was the highest at John F. Kennedy International Airport (JFK).

As was the case for airports, few significant differences in performance were found among the six security companies represented. No significant difference in d′ was exhibited across security companies and those sampled may, therefore, be considered equivalent in terms of IED detection performance.

Screeners were asked to indicate how confident they were in their judgments and screeners more often correctly judged bags not to contain an IED when they reported being very sure of their decision. However, this was not the case when deciding bags contained an IED. Screeners were no more likely to correctly report a bomb when very sure than when reporting that they were not so sure.

1. Introduction

1.1 Purpose

Threat objects such as weapons have historically resulted in the highest detection rates in operational settings. In contrast, detection rates for improvised explosive devices (IEDs) have been comparatively lower. IED detection performance was, therefore, of specific concern to this study and testing focused on the ability of screeners to detect IEDs.

The Federal Aviation Administration (FAA), in cooperation with the U.S. aviation industry, is developing new equipment and procedures to improve the national aviation security system. The FAA plans to operationally demonstrate some of these new technologies and procedures at diverse locations to verify their performance enhancement benefits. Before this demonstration takes place, however, a baseline of current aviation security performance must be established against which any performance improvements can be compared.

The purpose of this Test and Evaluation Report (TER) is to present, explain, and discuss the conduct and results of the Explosive Device Detection Baseline (EDDB) study that was conducted at the category X (CAT X) airports. The study was conducted in support of the Aviation Security Human Factors Program at the FAA Technical Center, Atlantic City International Airport, New Jersey, under Research Project Description #127 in support of Mission Need Statement #163.

1.2 Scope

The EDDB study evaluated the capability of airport baggage screeners to detect IEDs in carry-on (CO) passenger baggage using simulated black/white images. The screening environment was simulated by having operational airport baggage screeners at each of the 19 CAT X airports scan for IEDs in X-ray images presented by a computer-based library of digitized images.

1.3 Background

The FAA has had the responsibility of ensuring the safety of air travel since its creation in 1958. Both airports and aircraft are vulnerable to terrorist attacks, but airports pose a particular challenge to security. They must be readily accessible to the public, yet prevent persons with malicious intent from penetrating secure areas. Airport baggage screening is part of the FAA's security concept for airports in which a system comprising trained personnel, properly maintained and calibrated equipment, and appropriate procedures provides multiple layers of security from the airport perimeter to the aircraft door. Preboard baggage screening is part of the airport security checkpoint system and part of the total airport security system.

The threat to civil aviation security has changed dramatically in the last decade. This change in threat has resulted in new challenges for passenger and baggage screening. In the 1980s, the threat was hijacking. The role of the FAA in aviation security against hijacking was greatly expanded, especially after the 1985 hijacking of Trans World Airlines flight 847 in the Middle East. In the 1990s, there has been a shift from threats due to hijacking to concern about sabotage by bombing.

Improvements in technology available to hostile elements, especially in the area of explosive devices, have resulted in increased airliner vulnerability to bombing. Terrorists are reducing their use of prefabricated explosive devices, such as grenades, and opting for less detectable IEDs. An IED can be made from a variety of materials that may resemble innocent or everyday objects, such as batteries, wires, and digital clocks. For example, plastic explosives made with Semtex and C-4 can be shaped and molded into sheets or cubes that, when passed through X-ray screening devices, appear as innocent items such as books or radios. Terrorists have also learned to embed IEDs in electronic devices, as in the Pan American flight 103 disaster, making detection even more difficult. In addition, miniaturization and digitization of timing devices compounds the problem of IED detection with X-ray screening.

Sophisticated terrorists have the knowledge and materials to build difficult-to-detect IEDs. The potential for complete aircraft destruction, with the loss of hundreds of lives and the disruption of the National Airspace System, has increased. As a result of this shift toward a higher potential for disaster, the focus of civil aviation security has changed from hijackings to bombings. This shift has markedly increased the need for improvements in screener systems and operator training. The Associate Administrator for Civil Aviation Security and the Office of Civil Aviation Security Policy and Planning have identified the need for research into the performance of explosive device detection systems, particularly the human component, in detecting IEDs.

1.4 System Description

IED detection testing was conducted using computer-based black/white X-ray images. The test consisted of presenting 200 digitized bag images, 25 of which contained simulated IEDs built from modular bomb sets.

1.5 Critical Operational Issues and Criteria

The following are the Critical Operational Issues and Criteria (COIC) that were assessed during this study.

1.5.1 Issue 1 - Baseline Improvised Explosive Device Detection Performance

What is the baseline IED detection performance at each U.S. CAT X airport?

This issue was assessed by administering the same computerized IED detection test to samples of screeners from 19 CAT X airports and recording the number of hits (N_h) and the number of false alarms (N_{fa}) observed for each screener. Data analysis was then based on N_h and N_{fa} scores to calculate the probability of a correct detection (P_d) and the probability of a false alarm (P_{fa}). This was done for each screener as well as for the overall percentage correct and false alarm rates for each airport and security company.

Measures such as P_d and P_{fa}, conventionally used in detection studies, unfortunately vary considerably with changes in a screener's decision strategy. Their usage can thus give a distorted picture of screening efficacy. For example, a very liberal decision strategy (one requiring little convincing) can result in a very high P_d but does so at the cost of a P_{fa} that is also very high. Taken to its extreme, always saying an IED is present results in no misses but many false alarms. Signal Detection Theory (SDT) provides an approach for distinguishing between accuracy and decision strategy in evaluating screener performance in the IED detection task. The principal advantage over conventional P_d measures is that SDT provides a simple quantitative means of analyzing the two aspects of detection performance - the effectiveness with which IED vs. no IED discriminations are made (d′) and the decision criteria (c) used to select between the IED and no IED alternatives. This study emphasizes SDT in evaluating the two facets of detection. A discussion of the SDT paradigm is included in Appendix A.

Criterion None. This issue was investigative in nature.

MOP 1-1. The N_h observed for simulated IEDs.

MOP 1-2. The N_{fa} observed.

MOP 1-3. d′

MOP 1-4. c

MOP 1-5. P_d

MOP 1-6. P_{fa}

1.5.2 Issue 2 - Improvised Explosive Device Detection Performance Differences Across Airports

Are there significant differences in IED detection among CAT X airports?

This issue was assessed by determining any statistically significant differences in IED detection across the screener samples drawn from the 19 CAT X airports for issue 1. The IED detection performance values of d′, c, P_d, and P_{fa}, were compared across airports.

Criterion. None. This issue was investigative in nature.

MOP 2-1. See MOP 1-3.

MOP 2-2. See MOP 1-4.

MOP 2-3. See MOP 1-5.

MOP 2-4. See MOP 1-6.

1.5.3 Issue 3 - Improvised Explosive Device Detection Performance Differences Across Airport Security Companies

Are there significant differences in IED detection among airport security companies at the CAT X airports?

This issue was assessed by determining any statistically significant differences in IED detection across three large security companies (Argenbright, ITS, and Globe) and one group of smaller companies (Andy Frain, ATS, and ARC)

Criterion. None. This issue was investigative in nature.

MOP 3-1. See MOP 1-3.

MOP 3-2. See MOP 1-4.

MOP 3-3. See MOP 1-5.

MOP 3-4. See MOP 1-6.

1.6 Test and Evaluation Limitations and Impact

The computerized testing lacked some aspects of operational representativeness compared to actual screening duties at a security checkpoint. The full effects of this diminished representativeness are unknown but the test situation is expected to have resulted in better than usual performance.

Performance also may have been affected by the higher-than-normal threat presentation rate. Conditions that alter arousal level and/or attention to the task may affect c. However, the screeners were not given feedback on their accuracy or actual numbers of threat objects, so the effects are likely to be minimized.

1.7 Testing Milestones

Table 1 shows the milestones used for planning and reporting the test and evaluation process.

Table 1

Test and Evaluation Milestones

Milestone	Date	Responsible Organization
Test Concept/Design Approval	8 Sep 95	AAR-510
TEP Submitted	13 Jul 95	Contractor
Coordinate Airport Sites	Continuous	AAR-510
Test Readiness Review	8 Sep 95	AAR-510
TEP Approved	15 Sep 95	AAR-510
Testing Initiated	18 Sep 95	AAR-510, Contractor
Testing Completed	21 Nov 95	AAR-510, Contractor
TER Submitted	8 Dec 95	Contractor
TER Approved	29 Dec 95	AAR-510

2. Test Description

2.1 Subjects

The baseline sample consisted of 521 certified screeners: 19 screeners were from Luis Munoz Marin International Airport (San Juan) Airport (SJU), 26 were from Baltimore-Washington International Airport (BWI), and 28 were from each of the remaining 17 CAT X airports. This sample of screeners was drawn according to the relative market share of certified X-ray screeners for each security company across the CAT X airports. That is, if Company *X* employed 20% of the screeners working at the CAT X airports, 20% of the screeners in the sample were from Company *X*. However, only companies with 5% or more of the market share were included, except for ATS which was the only security company at the Lambert St. Louis International Airport (STL). The sample is further described in Table 2. The screeners completed an informed consent form (Appendix B), a personal information questionnaire (Appendix C), and received test protocol instructions (Appendix D) before participating in each test of the study. They were then given a vision test and the IED detection test.

Table 2

Screeners (N = 521) Sampled by Airport Security Company

Security Co.	Screeners	
	Airport	n
Andy Frain	Seattle Tacoma International Airport (SEA)	28
ARC	Atlanta William B. Hartsfield International Airport (ATL)	28
Argenbright	Washington National Airport (DCA)	28
	Denver International Airport (DIA)	28
	Washington Dulles International Airport (IAD)	28
	John F. Kennedy International Airport (JFK)	28
	Orlando International Airport (MCO)	28
ATS	Lambert St. Louis International Airport (STL)	28
Globe	Dallas/Ft. Worth International Airport (DFW)	28
	Houston Intercontinental Airport (IAH)	28
	Luis Munoz Marin International Airport (SJU)	19
ITS	General Edward Lawrence Logan International Airport (BOS)	28
	Baltimore-Washington International Airport (BWI)	26
	Detroit Metropolitan Wayne County Airport (DTW)	28
	Honolulu International Airport (HNL)	28
	Los Angeles International Airport (LAX)	28
	Miami International Airport (MIA)	28
	O'Hare International Airport (ORD)	28
	San Francisco International Airport (SFO)	28

2.2 Test Organization

Three test administrators were required for this study, one from the FAA and two from Galaxy Scientific Corporation. The test administrators shared various duties, which are detailed in Appendix E.

2.3 Test Organization Training

The test organization received training on all tests, procedures, and protocols before conducting the study.

2.4 Test Procedures

2.4.1 Pilot Study

The pilot study conducted for the Screener Proficiency Evaluation and Reporting System (SPEARS) Threat Image Projection study used the same IED detection testing equipment and protocol required for this study (Fobes & McAnulty, 1995). That effort also served as the pilot for this investigation.

2.4.2 Test Protocol

The testing activities at each CAT X airport required approximately 12 hours to complete. The test schedule consisted of seven 1.5-hour test cycles with a maximum of four screeners tested per cycle. The testing was conducted at two airports in relative geographic proximity to each other (for example, Chicago and Detroit) during a given workweek.

Each cycle proceeded as follows (the times are only examples). Screeners received a briefing and completed an informed consent form and a personal information questionnaire from 0800 to 0815. Vision testing (see Appendix F) took place from 0815 to 0830. IED testing took place from 0830 to 0915 and screener debriefing took place from 0915 to 0930. Screeners were informed that they could take a rest break at any time during the testing.

The IED testing took place on four computer-based image presentation devices (one device per screener). Before each testing cycle, the test administrator assigned each screener an identification number and ensured that each test device was serviceable and in the proper configuration. Before each test, screeners were briefed using the instructions contained in Appendix D.

2.4.2.1 Vision Testing

All screeners were given the Regan High Contrast Acuity Test to ensure that their vision was within the normal range.

2.4.2.2 Improvised Explosive Device Testing

Each screener was informed that IEDs would be present in some of the bag images presented during the test but were unaware of the test bag insertion order. Screeners carried out normal screening operations and did not receive operational direction from the test administrators. Each computerized test involved the presentation of 25 black/white X-ray CO test bag images containing simulated IEDs within a total series of 200 black/white X-ray CO bag images. All screeners in the study viewed the same 200 test and comparison bag images. To control for presentation order, seven randomly generated presentation orders for the 200 bag images were used.

The test administrator started the test trial and asked the screener to begin the test. The testing device automatically displayed the first bag image to the screener. The screener's task was to indicate whether the bag contained a threat image by pressing the appropriate key on the keyboard. Screeners were given 10 seconds to respond yes or no. At 11 sites, the screeners were aurally and visually prompted after 6 seconds; at 8 sites, the screeners received only a visual prompt. At the end of the 10 seconds, the computer-based test directed the screener to make an immediate response. After responding, the screeners were asked to indicate how confident they were in their response concerning the presence or absence of an IED in the bag image. The confidence rating choices were "very sure" and "not so sure." In combination with their prior yes or no judgment, the following categories resulted:

Category 1 - Yes, very sure
Category 2 - Yes, not so sure
Category 3 - No, not so sure
Category 4 - No, very sure

Screeners were allowed an additional 5 seconds to make this confidence response. If no response was given, the screener was aurally and visually prompted. The testing device recorded the response and automatically forwarded to the next image. The screeners received no feedback on the correctness of their responses.

2.5 Data

2.5.1 Improvised Explosive Device Detection Test

The N_h and N_{fa} was collected for each screener.

2.5.2 Vision Test

Visual acuity scores were collected for each screener using the Regan High Contrast Acuity Chart.

2.6 Database Management

Table 3 shows the database layout for the data collected for statistical analyses required to support the evaluation of the baseline data against the COIC.

All data will be retained for 5 years by the principal investigator for the project. In accordance with professional and ethical standards, the principal investigator will maintain separate records of performance data and the names of participating screening personnel.

Table 3

Screener Baseline Database

Screener ID
Airport
Security Company
Screener Characteristics
Experience
X-ray Equipment Experience
Date
Vision Test Score
IED Detection/Confidence Rating Scores
N_h
Bag Number for Each Hit
N_{fa}
Bag Number for Each False Alarm
d′
c
P_d
P_{fa}

2.7 Data Analyses

Data were described using descriptive statistics and analyzed using parametric statistics (analysis of variance [ANOVA], and Duncan post hoc test). Sections 2.7.1 through 2.7.4 describe the specific analyses performed.

2.7.1 Individual and Group Baseline Performance Data

The values of N_h and N_{fa} were used to derive the IED detection performance variables of d′, c, P_d, and P_{fa} for each screener.

The N_h was used to calculate the P_d as follows: $P_d = N_h/N_t$; where N_t was the number of test bag images presented to the screener.

The N_{fa} was used to calculate the P_{fa} as follows: $P_{fa} = N_{fa}/N_c$; where N_c was the number of comparison bags presented.

The above probabilities are used to determine SDT measures by converting P_d and P_{fa} to z-scores, where z-scores are abscissa values from the standard normal curve. Screener sensitivity was determined according to the formula: $d' = (z_{fa}) - (z_h)$; where z_{fa} and z_h were the z-score conversions of N_{fa} and N_h.

Receiver operating characteristic (ROC) d′ values were calculated. However, a review of the confidence rating data showed that for the comparison bags, a number of subjects were predominantly choosing the "No IED, very sure" confidence rating. This was reflected in the relatively low P_{fa} rates exhibited by the screeners. One of the assumptions of using confidence ratings to determine d′ is that the threat and comparison bags are fairly evenly distributed in terms of difficulty to analyze. Therefore, responses will be evenly distributed across confidence categories. This was not the case for the comparison bags set used in the computerized IED detection test, as reflected in the skewed responses to the comparison bags by some screeners. These skewed responses caused the ROC d′ values calculated from these responses to be inordinately high as compared with d′ values calculated from simple yes-no responses. Therefore, calculation of d′ was performed according to a yes-no response paradigm. The P_d and P_{fa} were used to determine screener response criterion as follows: $\underline{c} = .5(z_{fa} + z_h)$.

Mean (M) and standard deviation (SD) values were calculated for each airport and screening company for each measure.

2.7.2 Improvised Explosive Device Detection Performance Compared Across Airports

The second issue required the determination of any significant differences in IED detection performance across airports.

Four separate, single-factor, between-subjects designs were used to address this issue, as depicted in Table 4. The independent variable was airport (19 CAT X airports) and the dependent variables were the values of d′, c, P_d, or P_{fa} determined for issue 1. A one-way ANOVA was conducted across airports for each of these four dependent variables.

Table 4

Experimental Design for Airports

Airport (U.S. Category X)								
Airport 1 (n = 19)		Airport 2 (n = 26)		Airport n (n = 28)		. . .	Airport 19 (n = 28)	
s1	s15	s20	s34	s46	s60	. . .	s494	s508
s2	s16	s21	s35	s47	s61	. . .	s495	s509
s3	s17	s22	s36	s48	s62	. . .	s496	s510
s4	s18	s23	s37	s49	s63	. . .	s497	s511
s5	s19	s24	s38	s50	s64	. . .	s498	s512
s6	-	s25	s39	s51	s65	. . .	s499	s513
s7	-	s26	s40	s52	s66	. . .	s500	s514
s8	-	s27	s41	s53	s67	. . .	s501	s515
s9	-	s28	s42	s54	s68	. . .	s502	s516
s10	-	s29	s43	s55	s69	. . .	s503	s517
s11	-	s30	s44	s56	s70	. . .	s504	s518
s12	-	s31	s45	s57	s71	. . .	s505	s519
s13	-	s32	-	s58	s72	. . .	s506	s520
s14	-	s33	-	s59	s73	. . .	s507	s521
							N_T = 521	

2.7.3 Improvised Explosive Device Detection Performance Compared Across Airport Security Companies

The third issue required the determination of any significant differences in IED detection performance across security companies at the airports.

Four separate, single-factor, between-subjects designs were used to address this issue, as depicted in Table 5. The independent variable was airport security company (three large companies and one group of smaller companies combined) and the dependent variables were the values of d′, c, P_d, or P_{fa} determined for issue 1. A one-way ANOVA was conducted across security companies for each of these four dependent variables.

Table 5

Experimental Design for Airport Security Companies

Airport Security Companies							
Argenbright (n = 140)		ITS (n = 222)		Globe (n = 75)		Smaller Companies (n =84)	
s1	.	s141	.	s363	.	s438	.
s2	.	s142	.	s336	.	s439	.
s3	.	s143	.	s337	.	s440	.
s4	.	s144	.	s338	.	s441	.
s5	.	s145	.	s339	.	s442	.
s6	.	s146	.	s340	.	s443	.
s7	.	s147	.	s341	.	s444	.
s8	.	s148	.	s342	.	s445	.
s9	.	s149	.	s343	.	s446	.
s10	s140	s150	s362	s344	s437	s447	s521
						N_T = 521	

2.7.4 Accuracy as a Function of Confidence

The accuracy with which screeners were able to determine whether bags contained a bomb was also examined according to the confidence they had in their judgments. This was done by analysis of being correct differing as a function of reporting “Yes, Very Sure” (VS), “Yes, Not So Sure” (NS), “No, VS,” or “No, NS.”

3. Test And Evaluation Results

An ANOVA indicated that performance did not differ as a function of combined aural and visual versus only visual warnings (see section 2.4.2.2). This variable was accordingly not examined in the following analyses.

All of the following analyses were evaluated at the $p<.01$ level of confidence.

3.1 Issue 1 - Baseline Detection at Individual Airports

Mean performance variable values and standard deviations across CAT X airports, by individual airport and by airport security company are shown in Table 6.

Table 6

Performance Variable Means and Standard Deviations Across 19 Category X Airports, by Airport and by Security Company

	d′		c		P_d		P_{fa}	
	M	SD	M	SD	M	SD	M	SD
Across Airports	1.29	.46	-.46	.63	.55	.24	.17	.11
By Airport								
HNL	1.68	.61	-.40	.49	.64	.19	.12	.11
SEA	1.49	.38	-.15	.60	.68	.19	.21	.16
IAD	1.48	.41	-.44	.69	.60	.20	.16	.15
MCO	1.47	.40	-.55	.65	.58	.22	.14	.13
SFO	1.43	.50	-.65	.65	.52	.23	.12	.14
DFW	1.38	.47	-.49	.65	.56	.22	.15	.16
BWI	1.32	.45	-.47	.50	.51	.23	.13	.11
MIA	1.31	.39	-.35	.51	.63	.25	.25	.17
JFK	1.31	.48	-.06	.60	.68	.21	.29	.18
IAH	1.30	.36	-.64	.51	.51	.18	.13	.11
SJU	1.27	.41	-.51	.59	.47	.25	.13	.11
DTW	1.27	.36	-.89	.57	.41	.21	.09	.10
ORD	1.19	.49	-.59	.65	.46	.27	.14	.13
DIA	1.18	.40	-.49	.65	.51	.24	.19	.20
ATL	1.16	.45	-.37	.72	.57	.26	.22	.18
STL	1.14	.36	-.40	.65	.55	.23	.20	.18
BOS	1.13	.52	-.43	.60	.51	.28	.17	.13
DCA	1.12	.37	-.25	.48	.57	.22	.23	.17
LAX	1.00	.45	-.69	.66	.44	.24	.16	.12
By Security Company								
Argenbright	1.31	.41	-.36	.69	.57	.24	.21	.18
Globe	1.32	.41	-.49	.60	.56	.21	.15	.14
ITS	1.34	.50	-.54	.61	.52	.24	.15	.14
Smaller Companies	1.26	.40	-.30	.66	.60	.23	.21	.17

3.2 Issue 2 - Baseline Improvised Explosive Device Detection Performance Across Airports

The ANOVA resulted in significant differences across airports in screener d′, c, and P_{fa}. The following Duncan post hoc results were obtained:

Screener d′ ($F(18, 502) = 4.15$) was significantly higher at HNL than at BWI, ORD, DTW, DIA, ATL, STL, BOS, JFK, DCA, STL, ATL, IAH, MIA, and LAX. Screener d′ was significantly lower at LAX than at BWI, SEA, JFK, IAD, DFW, HNL, SFO, SJU, and MCO.

Screener c ($F(18, 502) = 2.91$) was significantly higher at JFK and SEA than at DTW.

Screener P_{fa} (F(18, 502) = 3.12) was significantly higher at JFK than at SFO, HNL, and DTW.

3.3 Issue 3 - Baseline Improvised Explosive Device Detection Performance Across Airport Security Companies

An ANOVA resulted in significant differences across security companies in screener P_{fa}. (F(3, 517) = 5.67). Duncan post hoc analyses indicated that the P_{fa} was significantly higher for the combined smaller security companies than for ITS.

3.4 Accuracy as a Function of Confidence

The accuracy with which screeners found bombs, under the two levels of their self-reported confidence, was examined with their conditional probabilities of a hit when VS [P(H | VS)] and when NS [P(H | NS)]. An ANOVA evaluated these probabilities of being correct as a function of screeners' response criteria (c) of liberal (-2.04 to -.765), neutral (-.764 to -.175), or conservative (-.174 to 1.305). No significant differences were found.

The accuracy with which screeners were able to determine bags to be free of bombs, under the two levels of self-reported confidence, was similarly examined. The conditional probabilities of a correct rejection (CR) when the screener was VS and when NS were compared under the above three ranges of criteria (c). The ANOVA indicated that the average P(CR | VS) of 0.85 was greater than that of 0.56 for P(CR | NS) [F(1,518) = 233.6].

4. Discussion

4.1 Issue 1 - Baseline Detection Performance at Individual Airports

Mean d′ was 1.29 across the airports, with an accompanying average P_d of 0.55 and average P_{fa} of 0.17. Clearly, there is room for improvement. Training systems (Fobes, et al, 1995b) will soon be tested to determine if they increase screener P_d while at the same time maintaining or reducing P_{fa} (that is, increase d′). This will allow improved detection of threats and, at the same time, avoid high false alarm rates that would reduce passenger flow.

4.2 Issue 2 - Baseline Improvised Explosive Device Detection Performance Across Airports

Except for a few statistically significant differences in screener d′, c, and P_{fa} between airports (those associated with highest and lowest scores), airports were generally equivalent on these performance variables. Further research to enhance screener performance may emphasize HNL and LAX, as HNL exhibited the highest mean d′ value and the Los Angeles International Airport exhibited the lowest mean d′.

4.3 Issue 3 - Baseline Improvised Explosive Device Detection Across Security Companies

No significant difference in d′ was exhibited across airport security companies and the security companies sampled may, therefore, be considered equivalent in terms of IED detection performance.

4.4 Accuracy as a Function of Confidence

Screeners more often correctly judged bags not to contain an IED when they reported being very sure of their decision. However, this was not the case when deciding bags contained an IED. Screeners were no more likely to correctly report a bomb when very sure than when reporting that they were not so sure.

5. Recommendations

As new technology and procedures for screener training are deployed and tested at the CAT X airports, it is recommended that screeners be tested at predetermined intervals to compare performance improvements against the baseline established by this study.

6. References

Coren, S. & Ward, L. M. (1989). Sensation and Perception (3rd ed.). Orlando, Florida: Harcourt, Brace, Jovanovich, Inc.

Fobes, J. L. & McAnulty, D. M. (1995). Test and evaluation plan for screener proficiency evaluation and reporting system (SPEARS) threat image projection. DOT/FAA/AR-95/50.

Fobes, J. L., McAnulty, D. M., & Klock, B. A. (1995a). Test and evaluation plan for the explosive device detection baseline (EDDB) study. DOT/FAA/AR-95/16.

Fobes, J. L., McAnulty, D. M., & Klock, B. A. (1995b). Aviation security human factors test and evaluation master plan for the airport demonstration. DOT/FAA/CT-95/30.

Green, D. M. & Swets, J. A. (1966). Signal detection theory and psychophysics. New York: Wiley.

Ingham, J. G. (1970). Individual differences in signal detection. Acta Psychologica, 34, 39-50.

Macmillan, N. A. & Creelman, C. D. (1990). Response bias: Characteristics of detection theory, threshold theory, and "nonparametric" indexes. Psychological Bulletin, 107, 401-413.

Snodgrass, J. G. & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. Journal of Experimental Psychology: General, 117, 34-50.

Wickens, C. D., (1992). Engineering psychology and human performance (2nd ed.). New York: HarperCollins Publishers, Inc.

APPENDIX A - SIGNAL DETECTION THEORY AND APPLICATION

The Signal Detection Theory (SDT) Paradigm

The explosive device detection task features screeners engaged in detection of an improvised explosive device (IED). SDT is a mathematical representation of human performance in deciding whether or not such a signal is present.

There are two response categories that represent a screener's detection: Yes (a modular bomb set [MBS] signal was present) or No (an MBS signal was not present). There are also two signal presentation states indicating that the MBS signal was present (signal) or absent (noise). A combination of screener responses and the signal state produces a 2 x 2 matrix (Figure A-1), generating four classes of operator responses, labeled hits, misses, false alarms, and correct rejections (Wickens, 1992).

		State of MBS Image	
		MBS Present	MBS Not Present
Screener Response	Yes	Hit	False Alarm
	No	Miss	Correct Rejection

FIGURE A-1. 2 X 2 MATRIX OF SCREENER RESPONSES AND STATE OF MBS IMAGE

a. A Hit is recorded when a baggage screener correctly detects an MBS in the scanned baggage.

b. A False Alarm is recorded when a baggage screener reports an MBS in the scanned baggage when none is present.

As indicated by Wickens (1992), the SDT paradigm assumes that operators perform two stages of information processing in all detection tasks: (1) sensory evidence is aggregated concerning the presence or absence of the signal, and (2) a decision is made about whether this evidence constitutes a signal. According to SDT, external stimuli generate neural activity in the brain. On the average, there will be more sensory or neural evidence in the brain when a signal is present than when it is absent. This neural evidence (X) increases in magnitude with stimulus (signal) intensity. If there is enough neural activity, X exceeds a critical threshold, X_c, and the operator decides "yes." If there is too little, the operator decides "no." Because the amount of energy in the signal is typically low, the average amount of X generated by signals in the environment is not much greater than the average generated when no signals are present (noise). Furthermore, the quantity of X varies continuously, even in the absence of a signal, because of random variations in the environment and the operator's level of neural firing (i.e., the neural "noise" in the operator's sensory channels and brain).

The relationship between the presence and absence of a signal can be seen in the hypothetical noise and signal plus noise distributions contained in figure A-2. The intersection of the two curves represents the location where the probability of a signal equals the probability of noise. The criterion value, X_c, chosen by the operator, is shown by the vertical line. All X values to the right ($X > X_c$) will cause the operator to respond "yes." All X values to the left generate "no" responses.

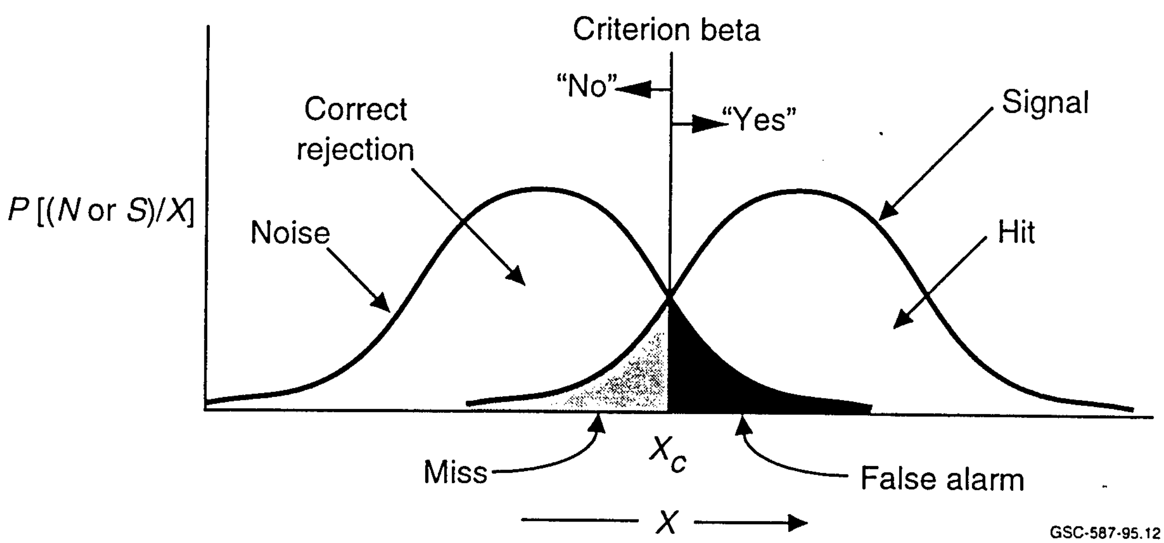

FIGURE A-2. HYPOTHETICAL SDT DISTRIBUTIONS (Wickens, 1992)

The different shaded areas represent the occurrences of hits, misses, false alarms, and correct rejections.

Procedures to Calculate SDT Probabilities

a. In SDT, the detection values are expressed as probabilities.

b. The probability of hit (P_h), miss (P_m), false alarm (P_{fa}), and correct rejection (P_{cr}) are determined by dividing the number of occurrences in a cell (figure A-1) by the total number of occurrences in a column.

c. The P_h (also referred as the probability of detection [P_d]) is calculated by dividing the number of IEDs detected (number of hits) by the total number of hits and misses: $P_m = 1 - P_d$.

d. The Pfa will be determined by the number of false alarms divided by the total number of false alarms and correct rejections: Pcr = 1 - Pfa.

Sensitivity (d′)

Sensitivity refers to the average amount of operator sensory activity generated by a given signal as compared with the average amount of noise-generated activity (Coren and Ward 1989). Baggage screeners may fail to detect (miss) an IED signal when employing a conservative response criterion. Correspondingly, the signal may be missed because the resolution of the detection process is low in discriminating signals from noise, even if the response criterion is neutral or risky.

d′ corresponds to the separation of the signal and noise distributions (figure A-2). As the magnitude of the signal increases, the mean of the signal distribution moves to the right. The proportion of signals detected (the P_d) changes as the distance between the signal and noise distributions varies. According to Wickens (1992), if the separation between the distributions is great, sensitivity is great, and an operator can readily distinguish between a signal plus noise event and a noise only event. Similarly, if the separation between signal and noise is small, d′ measures will be low, as will discriminability.

A representative table of Z values and ordinate values of the probability distribution related to hit and false alarm responses is needed to calculate d′. A complete table of the ordinate values of the standard normal distribution was used to calculate d′ for the Test and Evaluation Report.

Procedures to Calculate d′.

a. Find the false alarm rate from the outcome matrix in the HIT/FA column of the table.

b. Read across the table to the Z column (the label of the abscissa of the graph) and write it down.

c. Repeat these operations for the hit rate, calling the tabled value Z_h.

d. Calculate d′ using the following equation: $d' = Z_{fa} - Z_h$.

Operator Response Criterion

In any signal detection task, operator decision making may be described in terms of an operator response criterion. Operators may use "risky" response strategies by responding yes more often than no. A risky strategy allows operators to detect most of the signals that occur, but also produces many false alarms. Alternatively, operators may use "conservative" strategies, saying no most of the time, making few false alarms, but missing many of the signals.

Different circumstances may require conservative or risky strategies. For example, an appropriate IED detection strategy requires screeners to respond "yes" when there is a question regarding baggage contents. This response may produce false alarms when no threatening objects are present.

One recent parametric measure of response bias is c (Ingham, 1970; Macmillan & Creelman, 1990; Snodgrass & Corwin, 1988). The chief difference between the measure c and its parametric alternative β lies in the manner in which they locate the observer's criterion. Whereas the bias index β locates the observer's criterion by the ratio of the ordinates of the signal-plus -noise (SN) and noise (N) distributions, c locates the criterion by its distance from the intersection of the two distributions measured in z-score units. The intersection defines the point where bias is neutral, and the location of the criterion at that point yields a c value of 0. Conservative criteria yield positive c values, and liberal criteria produce negative c values. The measure c is computed as follows:

$$\underline{c} = .5(z_{fa} + z_h) \qquad (2)$$

APPENDIX B - INFORMED CONSENT

I, ____, have received a briefing by the FAA representative as to the purpose of the FAA study. I fully understand the purpose of the study and have been provided with the opportunity to ask questions of the FAA representative. The FAA representative informed me that the study will require a 30-minute briefing, a 1-hour and 10-minute performance test and a 20-minute vision test.

I understand that this study will impose very little stress. The only stress I may experience in this experiment may be some initial frustration as I learn how to use the testing system. As part of the data analysis, my data will be combined with that of other individuals and I will no longer be identifiable as a participant. I have been informed that my name will remain CONFIDENTIAL.

I have been informed that I have the right to withdraw from the experiment, and that the experiment monitor may terminate my participation in the interest of safety and the experiment. I also certify that I am at least 18 years of age.

I have been informed that if additional details are needed, I may contact any of the test administrators at the airport during the study, or contact James L. Fobes, Ph.D., (609) 485-4944, or Robert L. Malone, (609) 645-0900, upon completion of the study.

Signed: ____

Date: ______ / ______ /______

Witness: ____

Date: ______ / ______ /______

APPENDIX C - FAA IMPROVISED EXPLOSIVE DEVICE DETECTION BASELINE OPERATIONAL TEST AND EVALUATION SCREENER QUESTIONNAIRE

DATE:________________________ SUBJECT NUMBER:________________________

1. What is your sex?

   Male ______ Female _______

2. How long have you been a baggage screener?

   ____________Years ____________Months

3. How long have you been using X-ray equipment to screen baggage?

   ____________Years ____________Months

4. What is the highest education level that you have completed?

   8th Grade or Less _________

   Some High School _________

   High School Graduate _________

   Some College or University _________

   College or University Graduate _________

5. Do you wear glasses while using the X-ray equipment?

   Yes ________ No ________

6. Do you consider English to be your primary language?

   Yes ________ No ________

   If not, in which language are you most proficient? ________________________

   If not, are you proficient with English?

   Yes ________ No ________

APPENDIX D - IMPROVISED EXPLOSIVE DEVICE DETECTION TEST PROTOCOL

READ TO SCREENERS:

This is a test of how well X-ray machines work for detecting bombs. For this activity, we have put X-ray images of passenger bags in this computer. You will view the X-ray images, one bag at a time, and inspect each bag for a bomb. The images will be displayed on the video monitor.

For each image, the following question will be displayed on the monitor:

"Is there a bomb in this bag?"

If you see a bomb in the bag, respond "Yes," by pressing the "YES" key on the keyboard
If you do not see a bomb in the bag, respond "No," by pressing the "NO" key on the keyboard.

Since we want to know how well the X-ray machines work, it is important that you answer each question to the best of your ability. It is as important for you to say "no" when you do not see a bomb as it is to say "yes" when you do see a bomb.

For each X-ray image you will have up to 10 seconds to respond; however, you should answer as quickly as possible. You will hear a warning signal at <u>6 seconds to answer the question</u>. If you have not answered the question by then, you must do so immediately as you will only have 4 seconds left to give your answer. If you still have not responded after 10 seconds have elapsed, you will be reminded that you <u>must</u> give a yes or no answer.

Remember, you have up to 10 seconds to answer the question, but you must answer the question and you should answer it as quickly as possible.

After you have answered "yes" that there is a bomb in the bag or "no" there is not a bomb in the bag, you will be asked to further qualify your answer by indicating how confident you are in your answer. For each image, one of the following questions will be displayed on your monitor. If you answered yes to the previous question, the following question will appear on the monitor:

"How sure are you that there actually <u>was</u> a bomb in the bag?"

If you are very sure, respond by pressing "YES" on the keyboard. If you are not so sure press "NO" on the keyboard.

If you answered "no" to the previous question, the following question will now appear on the monitor:

"How sure are you that there <u>was not</u> a bomb in the bag?"

If you are very sure, respond by pressing "YES" on the keyboard. If you are not so sure press "NO" on the keyboard.

After you have either responded to the second question or 5 seconds have elapsed (whichever occurs first), a new image will appear on the monitor and you will answer the same questions all over again. This procedure will be repeated until you have viewed all X-ray images.

You will first take a practice test that consists of three questions. After the practice test, you will begin the real test.

Do you have any questions?

APPENDIX E - EXPLOSIVE DEVICE DETECTION BASELINE STUDY STATEMENT OF DUTIES FOR ADMINISTRATIVE PERSONNEL

Test Director:

Responsible for directing and overseeing all test activities and personnel.

Federal Aviation Administration

Test Manager:

Ensures screeners are greeted and completes required administration.
Conducts briefs and debriefs.
Manages daily test activities and responsible for starting and stopping test sequence as required.
Plans for and directs contingency activities.
Liaisons with security company administrative personnel.
Liaisons with airline personnel.
Ensures the presence of the proper screener in the required location at the required time.
Escorts screeners between security checkpoint and training and testing rooms.
Resolves any encountered problems with the test director.
Executes all required logistical activities as required.

Galaxy Scientific Corporation

IED Detection Tester

Ensures that the screener's number is recorded on test forms.
Ensures that test forms are given to Galaxy Scientific Corporation personnel.
Briefs and debriefs screeners regarding Improvised Explosive Device (IED) detection test.
Provides guidance and assistance to screeners on matters pertaining to conduct of test.
Administers IED detection test.

Federal Aviation Administration
Galaxy Scientific Corporation

Vision Tester

Ensures that the screener's number is recorded on test forms.
Ensures that test forms are given to Galaxy Scientific Corporation personnel.
Ensures that each screener completes a informed consent form and questionnaire.
Conducts visual acuity tests on screeners using the required test forms.

Federal Aviation Administration
Galaxy Scientific Corporation

APPENDIX F - REGAN HIGH CONTRAST ACUITY CHART

REGAN LOW CONTRAST ACUITY CHARTS

SCORE SHEET
SERIES 1

Chart A - 96% Contrast

Patient Name .... Date ....

Left Eye

Z	R	D	O	V	C	N	S	1
H	R	V	C	O	S	K	Z	2
N	D	C	O	H	R	V	S	3
K	V	R	Z	C	O	H	S	4
Z	N	V	K	D	S	O	R	5
D	C	R	V	H	N	Z	K	6
O	S	K	C	V	R	Z	N	7
S	N	H	K	C	D	V	O	8
N	R	D	C	O	K	S	Z	9
V	H	C	O	R	Z	D	N	10
H	R	O	S	C	V	K	N	11

Right Eye

Z	R	D	O	V	C	N	S	1
H	R	V	C	O	S	K	Z	2
N	D	C	O	H	R	V	S	3
K	V	R	Z	C	O	H	S	4
Z	N	V	K	D	S	O	R	5
D	C	R	V	H	N	Z	K	6
O	S	K	C	V	R	Z	N	7
S	N	H	K	C	D	V	O	8
N	R	D	C	O	K	S	Z	9
V	H	C	O	R	Z	D	N	10
H	R	O	S	C	V	K	N	11

Left Eye

Number of Errors	Line Number	Score

Right Eye

Number of Errors	Line Number	Score

It is important to urge the patient to guess each letter, even when uncertain.
Mark each error by crossing out each letter missed.